

J. Clin. Chem. Clin. Biochem.
Vol. 19, 1981, pp. 1145–1151

The Inadequacy of Normal Distribution Models for the Establishment of Assigned Values in Control Sera¹⁾

The Establishment of Assigned Values in Control Sera, II

By H. Passing

Hoechst AG, POB 800320, D-6230 Frankfurt (M) 80, Federal Republic of Germany

(Received February 23/July 15, 1981)

Summary: We compare 4 statistical models for the establishment of assigned values in a control serum which are based on the assumption of a normal distribution. The first model results in $\bar{x} \pm 2s$, whereas each of the following 3 models are based on a special analysis of variance. We studied by means of appropriate statistical tests the distributional properties of the data of the study described in part 1 of this series (Passing, H. et al. (1981) this j. 19, 1137–1144).

Many model assumptions are violated: The totality of data of each method was never normally distributed, normal distribution within laboratories was not given in 27 out of 67 cases, and precision and accuracy varied from reference laboratory to reference laboratory. Moreover, assigned values and uncertainty intervals calculated by means of these methods can be misleading to the customer. Therefore, these models cannot be applied, and a distribution-free procedure has to be used instead.

Die Unangemessenheit von Normalverteilungsmodellen bei der Ermittlung von Sollwerten in Kontrollseren
Ermittlung von Sollwerten in Kontrollseren, II

Zusammenfassung: Wir vergleichen 4 statistische Modelle zur Ermittlung von Sollwerten in einem Kontrollserum, die auf der Annahme einer Normalverteilung basieren. Das erste Modell führt zu $\bar{x} \pm 2s$, während die folgenden drei Modelle jeweils auf einer speziellen Varianzanalyse beruhen. Mit Hilfe geeigneter statistischer Tests untersuchten wir die Verteilungseigenschaften der Daten der in Teil 1 dieser Reihe (Passing, H. et al. (1981) this j. 19, 1137–1144) beschriebenen Studie.

Viele Modellannahmen sind verletzt: Die Gesamtheit der Daten jeder Methode war niemals normalverteilt, in 27 von 67 Fällen war keine Normalverteilung innerhalb der Laboratorien gegeben, und Präzision und Richtigkeit variierten von Referenzlabor zu Referenzlabor. Außerdem können Sollwerte und Sollbereiche, die nach diesen Methoden berechnet werden, für den Anwender irreführend sein. Deshalb können diese Modelle nicht angewandt werden, und an ihrer Stelle muß ein verteilungsfreies Verfahren benutzt werden.

Introduction

Various models for the establishment of assigned values and their uncertainty intervals in a control serum are known from the literature (3, 4, 5). They can be discriminated according to the following two points of view.

First. Some models (4, 5) postulate that the analytical values have certain special distributional properties. Such an assumption is then made *a priori* and will in general not be checked again (probably it is not checked

even when making the assumption). The consequences of its violation are usually not known even though they may be serious. Conversely, it is possible to define procedures which are not based on such assumptions (3) so that they are applicable more generally.

In this paper we point out the mathematical assumptions of four different models, and we examine by means of appropriate statistical tests whether they are justified. Thus we are able to assess the applicability of the models in l. c. (4, 5).

Secondly. There are two different philosophies concerning the meaning of an uncertainty interval. In l. c.

¹⁾ A preliminary report is given in l. c. (1)

(5) it is, mathematically speaking, a confidence interval i.e. it is very likely that the correct value of the control serum due to the method applied will be within this interval. Therefore, the width of the interval is a measure of the precision of the computed assigned value and not of single analytical values. In l.c. (3, 4), however, the uncertainty intervals are prediction intervals i.e. it is very likely that one additional analytical value established by the same method will be within it if systematical deviations do not occur.

Since assigned values of control sera and their uncertainty intervals are used for the internal quality control, i.e. for monitoring analytical values, it is clear that prediction intervals are more appropriate for this purpose than confidence intervals. We therefore confine ourselves to prediction intervals.

Materials and Methods

We used the data of the study performed by laboratories of members of the Verband der Diagnostica- und Diagnostica-geräte-Hersteller (VDGH, Association of Diagnostics and Diagnostics Instrumentation Manufacturers) as described in l.c. (2). Briefly, it was based on the constituents creatinine, glucose, urea, alanine and aspartate aminotransferase², creatine kinase², and γ -glutamyltransferase². For each constituent we had 9 to 11 laboratories (see l.c. (2) or table 4 below) each having performed double determinations in approximately 18 independent series of the unknown sample, of the known control, and of a blind control, respectively.

Two cases were discriminated generally:

Case 1

Both values of the double determinations of the unknown sample were used. They were accepted as valid in a series if both values of the known control as well as both values of the blind control of this series were within the uncertainty limits of the respective assigned values given by the manufacturer(s), cf. l.c. (2). In all other cases the whole series was eliminated.

Case 2

The first values only of the double determinations of the unknown sample were used simulating single determinations, cf. l.c. (2). They were accepted as valid in a series if the first value of the known control as well as the first value of the blind control of this series were within the uncertainty limits of the respective assigned value. In all other cases the whole series was eliminated.

We considered the following statistical models for the unknown sample, for details thereof see l.c. (6).

2) Enzymes:

Glutamate-pyruvate-transaminase = alanine aminotransferase = L-alanine: 2-oxoglutarate aminotransferase EC 2.6.1.2;

glutamate-oxalacetate-transaminase = aspartate aminotransferase = L-aspartate: 2-oxoglutarate aminotransferase EC 2.6.1.1;

γ -glutamyl-transferase = (5-glutamyl)-peptide: aminoacid 5-glutamyltransferase EC 2.3.2.2;
creatinine kinase = ATP: creatine N-phosphotransferase EC 2.7.3.2.

Model 1

The model equation for the analytical values is

$$x_k = \mu + e_k$$

where k runs over all available values. μ is the general mean. The first assumption (A_1) concerns the errors e_k .

(A_1): all e_k come from one common normal distribution, which implies equal accuracy. Furthermore, (A_1) implies that the variance of the e_k is the same for all laboratories, namely σ_1^2 , which means equal precision within laboratories.

Moreover, all analytical values must be independent.

Model 2

The model equation for the analytical values is

$$x_{ik} = \mu + l_i + e_{ik}$$

where i enumerates the laboratories and k all valid values within the i -th laboratory. μ is the general mean. l_i is the deviation of the i -th laboratory from μ thus permitting various accuracy within laboratories, assumed as a fixed effect. Regarding the errors e_{ik} it is assumed that

(A_2): the e_{ik} of all laboratories come from one common normal distribution,

(A_3): the variance of the e_{ik} is the same for all laboratories, namely σ_2^2 , which means equal precision within laboratories.

All analytical values must be independent.

This model is contained in l.c. (4). It should be stressed that in l.c. (5) the same assumptions must be made, though both models are quite different.

Model 3

The model equation for the analytical values is

$$x_{ik} = \mu + L_i + e_{ik}$$

where i , k , and μ have the same meaning as in model 2. L_i is the deviation of the i -th laboratory from μ , as a random effect. This means that the laboratories included are considered as random samples out of a population. It is assumed that

(A_4): all L_i come from a normal distribution with variance σ_L^2 .

The assumption concerning the errors e_{ik} is that

(A_5): the e_{ik} of all laboratories come from one common normal distribution.

This implies that

(A_5'): the variance of the e_{ik} is the same for all laboratories, namely σ_e^2 , which means equal precision within laboratories.

Additionally, the L_i and the e_{ik} are all assumed to be independent so that the variance of the analytical values is $\sigma_3^2 = \sigma_L^2 + \sigma_e^2$.

This model is also contained in l.c. (4).

Models 1 to 3 need not be modified, no matter if case 1 or case 2 occurs, but we must modify model 4. Additionally it is necessary with model 4 to restrict ourselves to a fixed number of valid series in common to all laboratories, i.e. to have a balanced design.

Model 4.1

The model equation for the analytical values is

$$x_{ijk} = \mu + L_i + S_j + (LS)_{ij} + e_{ijk}$$

where i enumerates the laboratories, j the series within each laboratory and k the two values of the double determination. μ is again the general mean. L_i is the random deviation of the i -th laboratory from μ . The laboratories included are again considered as random samples out of a population. It is assumed that

(A_6): all L_i come from a normal distribution with variance σ_L^2 .

Analogously, S_j is the random deviation of the j -th series from μ . The series performed are considered as random samples out of a population. It is assumed that

(A₇): all S_j come from a normal distribution with variance σ_S^2 .

As to the random interaction term $(LS)_{ij}$ it is assumed that

(A₈): all $(LS)_{ij}$ come from a common normal distribution with variance σ_{LS}^2 .

Moreover, concerning the analytical errors e_{ijk} , it is postulated that

(A₉): the e_{ijk} of all laboratories and all series come from one common normal distribution.

This implies that

(A_{9'}): the variance of e_{ijk} is the same for all laboratories, namely σ_e^2 , which means equal precision within series for all laboratories.

Finally, the L_i , S_j , $(LS)_{ij}$, and e_{ijk} must all be mutually independent so that the variance of the analytical values is $\sigma_{4.1}^2 = \sigma_L^2 + \sigma_S^2 + \sigma_{LS}^2 + \sigma_e^2$.

Model 4.2

The model equation for the analytical values reduces to

$$x_{ij} = \mu + L_i + S_j + e_{ij}$$

where the symbols all have the same meaning as in case 1. The assumptions (A₆), (A₇), (A₉), and (A_{9'}) are made. Since L_i , S_j , and e_{ij} are postulated as independent the variance of the analytical values is $\sigma_{4.2}^2 = \sigma_L^2 + \sigma_S^2 + \sigma_e^2$.

The following holds for all four models: The assigned value is the total mean \bar{x} of all analytical values, and its uncertainty limits are $\bar{x} \pm 2s$ where s is the estimation of σ_1 , σ_2 , σ_3 , $\sigma_{4.1}$, $\sigma_{4.2}$, respectively.

The following statistical test procedures were all based on an error probability of $\alpha = 5\%$.

For testing (A₁), all analytical values x_k were pooled. The *Anderson-Darling* test due to *Pettitt* (7) was applied to test their common normality.

For testing (A_{1'}), the values x_k within each laboratory only were pooled. The variances within laboratories were compared with each other using the test of *Levene* (8).

For testing (A₂), we distinguished as follows. In case 1 the differences within series were tested for normality separately for each laboratory using l.c. (7). In case 2 however, the first values themselves were tested by means of l.c. (7) separately for each laboratory.

Assumption (A₃) is obviously identical to (A_{1'}).

For testing (A₄), the $\mu + L_i$ of the model equation were estimated by the mean of the first 15 valid series of analytical values of each laboratory (this restriction to 15 series was necessary here in order to avoid a systematic error in calculating these mean values caused by different numbers of valid series available per laboratory). These mean values were tested for normality using again l.c. (7).

For testing (A₅), the error terms e_{ik} were estimated by $x_{ik} - \bar{x}_i$ where \bar{x}_i denotes the mean of all valid series of the i -th laboratory. They were tested for common normality by means of l.c. (7).

For testing (A_{5'}), the above estimations of e_{ik} were used again. Thus, testing for homogeneity of variances of e_{ik} or of the x_{ik} is equivalent when using l.c. (8).

For testing (A₆) up to (A₉), the first 15 valid series only of each laboratory were used.

Testing for (A₆) is identical with testing for (A₄).

For testing (A₇), for each serial number j the $\mu + S_j$ of the model equation were estimated by the mean over all laboratories. These mean values were tested for normality by use of l.c. (7).

For testing (A₉) in case 1, the error term e_{ijk} was estimated by $x_{ijk} - \bar{x}_{ij}$ where $\bar{x}_{ij} = (x_{ij1} + x_{ij2})/2$. Because the estimations of e_{ij1} and e_{ij2} differ by their sign only we used only the independent $x_{ij1} - \bar{x}_{ij}$ in the test (7).

Assumptions (A₈) in case 1 and (A₉) in case 2 however could not be tested because there are no estimations available for $(LS)_{ij}$ and e_{ij} which are "sufficiently" independent. (A_{9'}) has not been tested, since variance estimations based on one or two values only are impossible or without sense.

Additionally, in model 4 the analysis of variance was performed in order to estimate the variance components σ_L^2 , σ_S^2 , σ_{LS}^2 , σ_e^2 . Concerning model 4.1 the hypothesis $\sigma_{LS}^2 = 0$ was tested as described in l.c. (6).

Results

Table 1 contains the minimal and maximal number of valid series per laboratory and constituent after performance of the above validity checks when both values (case 1) as well as first values only (case 2) of the double determinations are used.

A validity check based on both values is more restrictive than one based on first values only. That is why the figures of case 1 partly are smaller than those of case 2. Additionally, the table gives the total number of valid series per constituent for both cases.

At least 15 series per laboratory and constituent are valid. Taking exactly 15 valid series for testing (A₄) and (A₆) up to (A₉) is caused by this fact.

Table 2 shows the results of the statistical tests concerning the assumptions in case 1.

Table 3 gives the results of the statistical tests concerning the assumptions in case 2.

Tab. 1. Minimal and maximal number of valid series per laboratory and total number per constituent.

Constituent	Case 1 ^a)			Case 2 ^a)		
	Min.	Max.	Total ^b)	Min.	Max.	Total ^b)
Creatinine	15	17	149	16	17	152
Glucose	15	17	146	16	18	152
Urea	16	18	159	17	18	161
Alanine aminotransferase	15	17	179	16	17	184
Aspartate aminotransferase	17	19	197	17	21	199
Creatine kinase	16	17	150	16	17	151
γ -Glutamyl-transferase	15	18	157	15	21	160

^a) Using both values (case 1) or first values only (case 2) of the double determinations.

^b) Summed up over all laboratories.

Tab. 2. Results of statistical tests^{a)} using both values of double determinations (case 1).

Constituent	Assumption ^{a)}	(A ₁)	(A ₁) ^{b)}	(A ₄) ^{c)}	(A ₅)	(A ₇)	(A ₉)
Creatinine		* ^{d)}	*	n.s. ^{e)}	*	n.s.	*
Glucose		*	*	n.s.	*	n.s.	*
Urea		*	*	n.s.	*	n.s.	*
Alanine aminotransferase		*	*	n.s.	*	n.s.	*
Aspartate aminotransferase		*	*	n.s.	*	n.s.	*
Creatine kinase		*	*	n.s.	*	n.s.	*
γ-Glutamyltransferase		*	*	n.s.	*	n.s.	*

^{a)} For details see materials and methods.

^{b)} Also (A₃) and (A₅').

^{c)} Also (A₆).

^{d)} Significant violation ($\alpha = 5\%$) of assumption.

^{e)} Assumption not significantly violated ($\alpha = 5\%$).

Tab. 3. Results of statistical tests^{a)} using first values only of double determinations (case 2).

Constituent	Assumption ^{a)}	(A ₁)	(A ₁) ^{b)}	(A ₄) ^{c)}	(A ₅)	(A ₇)
Creatinine		*	*	n.s.	n.s.	n.s.
Glucose		*	*	n.s.	*	n.s.
Urea		*	*	n.s.	n.s.	n.s.
Alanine aminotransferase		n.s.	*	n.s.	n.s.	n.s.
Aspartate aminotransferase		*	*	n.s.	*	n.s.
Creatine kinase		*	*	n.s.	*	n.s.
γ-Glutamyltransferase		*	*	n.s.	*	n.s.

^{a)} For details see materials and methods.

^{b)} Also (A₃) and (A₅').

^{c)} Also (A₆).

Model 1

The tests are based on at least $2 \cdot 146 = 292$ values in case 1 or 151 values in case 2. The analytical values of all laboratories do not have a common normal distribution. This was also clear from the frequency distribution of the data themselves because the distribution was often skew or even not unimodal. Therefore, the assumption (A₁) postulating one common normal distribution of all analytical values will be violated in general.

As to assumption (A₁'), the variances of analytical values always differ significantly indicating different

precision within laboratories. In fact, the standard deviations between laboratories differed by factors 2.5 up to 8. These tests are based on at least 15 series each.

Consequently, the assumptions of model 1 are flagrantly violated so that it must be rejected. Therefore, the establishment of assigned values and the calculation of their uncertainty intervals cannot be based on the familiar formula $\bar{x} \pm 2s$.

Model 2

The results of the tests concerning (A₂) are given in table 4 which are based on at least 15 series each.

For each test performed there is theoretically a chance of $\alpha = 5\%$ of getting a significant result in spite of not being correct. Summing up over all constituents we find 27 or 13 significances out of a total of 67. If all significances were wrongly found the proportions 27/67 or 13/67 should not be statistically higher than the above 5%. This was tested by means of l. c. (9). Both proportions turned out to be statistically higher than

Tab. 4. Results of statistical tests concerning assumption (A₂)^{a)}.

Constituent	Laboratories ^{b)}	Both values used ^{c)}	First values used ^{c)}
Creatinine	9	0	0
Glucose	9	4	2
Urea	9	4	1
Alanine aminotransferase	11	4	2
Aspartate aminotransferase	11	7	4
Creatine kinase	9	5	0
γ-Glutamyltransferase	9	3	4
Total	67	27	13

^{a)} For details see materials and methods.

^{b)} Number of laboratories included in this study.

^{c)} Number of laboratories thereof with distribution of analytical values deviating significantly ($\alpha = 5\%$) from normality.

5%. Therefore the tested values need not be distributed normally in general. This is true for the analytical values themselves in case 2.

In case 1 however the differences within the double determinations were tested. They turn out not to be distributed normally in general. It can be derived that the analytical values are not distributed normally in general, cf. l. c. (10).

In fact, the distribution of analytical values often looked skew within laboratories.

From this remarkable result it can be concluded that (A₂) will be violated. The results of our method of testing (A₅), based on a straightforward reasoning, confirm this statement.

Assumption (A₃) is always statistically violated. From the results of testing (A₂) and (A₃) it must be con-

cluded that model 2 covering equal precision within laboratories and differing accuracy is also inappropriate. The same statement holds for the first model of l.c. (4) and that of l.c. (5).

Model 3

No constituent shows significant deviation from normality of the $\mu + L_i$ in any case. These tests are based on 9 up to 11 laboratories each. We tested the $\mu + L_i$ because of independence but not the L_i themselves. The result naturally applies to the deviations L_i of accuracy. Therefore the distributional aspect of (A_4) formally seems not to be problematical. But basically it is assumed that the laboratories used for evaluation are a random sample out of the population (assumed large) of all laboratories performing this method. In practice however it occurs that only few reference laboratories are available for certain methods. So, the statistical result may be an answer to a non-realistic question.

Assumption (A_5) is very often statistically violated. The tests are based on at least 292 or 151 values, respectively. It follows that the errors e_{ik} of all laboratories will not have a common normal distribution in general – in spite of elimination of possible effects L_i of different accuracy! It might be objected here that the estimations of e_{ik} used are mutually dependent within laboratories so that the above testing procedure would not be correct. But this mutual dependence decreases with increasing number of values per laboratory. It should be negligible with at least 30 or 15 values per laboratory. In fact, most distributions of the e_{ik} turned out to be skew. These findings assure our result concerning (A_2).

Assumption (A_5') is always statistically violated. Therefore it must be concluded that model 3 covering equal precision of laboratories and randomly differing accuracy is inadequate. The same statement holds for the second model in l.c. (4).

Model 4

The results concerning (A_6) and their interpretation are identical with those of (A_4).

A somewhat different consideration concerns (A_7). Indeed, we never find a significant deviation from normality of the $\mu + S_j$ in any case based on 15 series each. We conclude that the random deviations S_j can have a normal distribution. The series performed are assumed as a random sample out of the large population of all series which are imaginable or might be performed. This postulate does not make sense because series are produced consecutively and not on a random basis. However, it ensures that analytical values of different series will be "more independent" than values of multiple determinations of the same series.

Assumption (A_9) in case 1 is always statistically violated. Each test is based on at least 135 values. So the analytical errors e_{ijk} within series of all laboratories and all series will not have a common normal distribution in general – in spite of elimination of possible effects of different accuracy between series. In fact, most distributions of analytical errors were skew.

The model seems acceptable statistically in case 2 but not in case 1. This contradiction is clarified by the following arguments. First, assumption (A_9) could not be tested in case 2. Furthermore, the absence of additivity of laboratorial and serial effects will be shown in both cases by using further mathematical arguments. Since the model equation of case 2 does not cover this possibility model 4.2 is inappropriate. Indeed, in case 1 the model equation does cover this possibility by means of the interaction term. But since assumption (A_9) is violated in case 1 it can be summarized: Models 4.1 and 4.2 must be refused. They cover equal precision within series and randomly differing accuracy between laboratories and series.

In detail, some variance components which are clearly expected to be positive are given in table 5.

Tab. 5. Some results of the analysis of variance of model 4^a).

Constituent	Signs ^b of s_S^2		F-value for interaction
	Case 1 ^c	Case 2 ^c	
Creatinine	> 0	< 0	12.4
Glucose	> 0	> 0	3.8
Urea	> 0	< 0	2.2
Alanine aminotransferase	> 0	< 0	2.9
Aspartate aminotransferase	> 0	< 0	5.3
Creatine kinase	> 0	< 0	2.8
γ -Glutamyltransferase	> 0	< 0	3.7

^a) For details see materials and methods.

^b) s_S^2 is the estimation of σ_S^2

^c) Using both values (case 1) or first values only (case 2) of the double determinations.

In case 2 most estimations s_S^2 of σ_S^2 are negative whereas s_L^2 was always positive. In case 1 however, all s_S^2 and all other variance components are positive. This is not caused by certain distributional assumptions being violated because it is a numerical and not a statistical effect. The reason is that the assumption of additivity of laboratorial and serial effects in the model equation cannot be correct in case 2. Unfortunately it is impossible to modify this assumption by enlarging the model in case 2. In case 1 however, non-additivity is explicitly contained in the model equation by means of the interaction term which causes all variance components to be positive.

The importance of this interaction term is stressed by the fact that the F-value for testing the hypothesis $\sigma_{LS}^2 = 0$, i.e. the absence of any interaction in case 1, is

always higher (cf. table 5) than a "critical value" of approximately 1.35. So interactions are "significantly" present, cf. l. c. (6). But this conclusion would be correct only if all model assumptions were fulfilled – in contrast to our result. Therefore it is necessary to avoid the notion of significance here and to state only qualitatively: The F-values found are much higher than 1.35 so that the existence of interactions is shown with high probability. s_S^2 was often found negative but never s_L^2 . The reason is that for each constituent s_L^2 was much larger than the absolute value of s_S^2 . Therefore s_L^2 was less sensitive to become negative than s_S^2 . This reflects the well known fact that the variance between laboratories usually is larger than between series.

Discussion

The assumptions of all four models can be divided into two classes:

Assumptions (A_4), (A_6), (A_7), and (A_8) concern properties of effects admitted by the respective model. Partly they restrict the applicability of the models from practical considerations.

The other assumptions however concerning the distribution of errors imply that these errors are symmetrically distributed. Actually we found most distributions skew, especially the distribution of analytical errors within series (model 4). From this the skewness of the error distributions of models 1 and 3 found when examining (A_1) and (A_5) can be clarified: It cannot completely be explained by deviation of single laboratories or even series which are not covered by the respective model.

Furthermore, our results confirm the well known fact that precision differs between laboratories.

Now we must stress the consequences of a violation of assumptions.

In case of a skew error distribution it is no longer meaningful to give \bar{x} as assigned value because many customers expect that approximately 50% of their analytical values are on each side of the given assigned value. They would then be falsely irritated or call in question the accuracy of the assigned value if they did not find this to be the case. In the case of symmetry however this problem cannot arise, since the mean value \bar{x} and median are then approximately identical. But by giving the symmetrical uncertainty limits $\bar{x} \pm 2s$ with s of the respective model the customer is misled into believing that symmetry exists. Therefore, skew uncertainty limits should be given in case of skew distributions in order to assist the customer rather than to confuse him.

There is still another serious statistical argument. Everybody believes that with 95% probability a further analytical value is expected to be within the uncertainty interval given if systematic deviations do not occur.

This statistical statement however depends essentially on all assumptions being met. For instance, it would not

suffice to have symmetry, but the stricter assumption of normality must be fulfilled. Conversely, nobody is able to assess the correct probability level if the assumptions are violated. In fact, it can be much less than 95%.

Indeed, we have analysed only certain constituents. But therefrom we can derive that assumptions of models 1 up to 4 can be violated in practice. Similar results are given in l. c. (3).

Apart from these results concerning distributions there is another fault of models 2, 3, and 4. Their model equations include differences of accuracy between the reference laboratories which is supported by our results.

At the moment of establishing assigned values this possibility is taken into consideration by means of the respective variance estimation, which only results in different widths of the respective uncertainty intervals.

Generally, the intervals of models 2, 3, or 4 are smaller than those given by model 1 where differences of accuracy are neglected. However, the assigned value itself is an estimation of the general mean μ , and so the uncertainty limits give a prediction interval for a further analytical value distributed around μ . The widths of these intervals do not include any differences of accuracy.

Yet, at the moment when a customer applies this control serum as internal quality control it must be conceded that his accuracy may differ from μ because models 2, 3, and 4 admit different accuracy even for the reference laboratories. In fact, its accuracy might be $\mu + l$ for a certain deviation l of accuracy so that analytical values are distributed around $\mu + l$. Then the given uncertainty limits are misleading. Rather, this laboratory would need uncertainty limits around $\mu + l$. More precisely, for each laboratory using this control serum special limits fitted to the respective $\mu + l$ would be necessary.

This is completely impossible, firstly from practical considerations, and secondly because the individual $\mu + l$ cannot be known in advance. Consequently, from this point of view models 2 up to 4 are inappropriate.

Therefore, other models should be defined in order to have a strategy for the evaluation of assigned values and their uncertainty limits which is applicable to all constituents and which is always correct. Such models do exist.

All models based on an analysis of variance and so using \bar{x} and some kind of s need assumptions concerning normality and equal precision. Necessarily they cover differences of accuracy of the reference laboratories only. Therefore this complete class of models will be inadequate. The difficulties described can be avoided by using distribution-free techniques. In part 3 of this series (11) we compare three distribution-free procedures.

Acknowledgements

This study was initiated by the Committee of Clinical Chemistry of the VDGH.

The author thanks *B. Müller* (Behringwerke AG) for many hints, and for assistance in writing this paper and to *W. Bablok* and *M. Glocke* (Boehringer Mannheim GmbH) for their co-operation in solving statistical problems. The determinations of the cited constituents were performed by *G. von Seydlitz* (Asid

Bonz und Sohn GmbH), *A. Berke* (Boehringer Ingelheim Diagnostika GmbH), *A. Benozzi*, *F. Bosslet*, *H. Brettschneider* (all Boehringer Mannheim GmbH), *W. von Thun* (Goedecke AG), *H.-G. Eisenwiener* (F. Hoffmann – La Roche & Co AG), *B. Carl* (Dr. Bruno Lange GmbH), *F. Schindler* (E. Merck), *R. Spaethe* (Merz und Dade, Abteilung der American Hospital Supply Deutschland GmbH), and *D. Schlitzer* (Technicon GmbH). The consequences from the results were drawn jointly by all participants.

References

1. Passing, H., Glocke, M., Brettschneider, H. & Müller, B. (1980) *Lab. Med.* **4**, 154–159.
2. Passing, H., Müller, B. & Brettschneider, H. (1981) *J. Clin. Chem. Clin. Biochem.* **19**, 1137–1144.
3. Hansert, E. & Stamm, D. (1980) *J. Clin. Chem. Clin. Biochem.* **18**, 461–490.
4. Schumann, V. (1976) *Med. Labor* **29**, 271–276.
5. National Committee for Clinical Laboratory Standards (NCCLS), Calibration reference materials and control materials in clinical chemistry (1975).
6. Scheffé, H. (1959) *The analysis of variance*. Wiley, New York.
7. Pettitt, A. N. (1977) *Appl. Statist.* **26**, 156–161.
8. Levene, H. (1960) Robust tests for equality of variances. In: *Contributions to probability and statistics*, (J. Olkin, ed.) Stanford, pp. 278–292.
9. Lehmann, E. L. (1970) *Testing statistical hypotheses*. Wiley, New York, pp. 128–129.
10. Loève, M. (1963) *Probability theory*. Van Nostrand, Princeton 1963, p. 271.
11. Passing, H. (1981) *J. Clin. Chem. Clin. Biochem.* **19**, 1153–1166.

For reprint requests:
Geschäftsstelle des
Verband der Diagnostica- und
Diagnosticgeräte-Hersteller (VDGH)
Karlstr. 21
D-6000 Frankfurt 1